图书情报工作
第66卷 第8期 2022年4月

LDA模型下我国公共图书馆微信平台阅读推广内容主题研究*

■ 李倩[1,2] 王帅[2]

[1] 山西大学经济与管理学院 太原 030006 [2] 中国科学技术信息研究所 北京 100038

摘 要：[目的/意义]识别国内公共图书馆借助微信公众平台开展阅读推广的内容主题及演化特征。[方法/过程]获取我国31家省级公共图书馆微信号发布的推文，对阅读推广内容进行界定和选取，运用LDA主题模型实现对阅读推广内容的主题识别、演化和阅读量分析。[结果/结论]发现当前公共图书馆借助微信开展的阅读推广内容主题可分为九大类，涵盖范围较广，每个主题随着时间的推移呈现不同的演化趋势，每个主题的阅读量也体现出不同的受欢迎程度。从推广内容的科学规划配比、推广内容从属地位的逆转、特殊人群推广内容设置、推广内容及形式的创新设计、打造品牌推广内容等角度对公共图书馆未来的微信阅读推广工作提出建议。

关键词：LDA模型 微信公众平台 阅读推广 主题挖掘

分类号：G250

DOI：10.13266/j.issn.0252-3116.2022.08.008

1 引言

图书馆拥有丰富的文献书籍资源，能够为读者提供各种资源获取服务，是传播知识、传递文化、传承文明的重要机构。公共图书馆积极有效地开展阅读推广工作有助于激发人们的阅读兴趣和阅读行为，进而养成全民阅读的良好习惯和阅读氛围。当前微信作为我国第一大社交媒体，在2021年的第三季度中其月活跃人数持续增长至12.62亿[1]，微信因其强大的交互功能已成为人们生活中不可缺少的社交APP。公共图书馆利用自媒体平台，将其高效、便捷、互动、共享性质应用到图书馆的推广交流中，这对阅读推广工作的深入发展具有不可替代的作用。公共图书馆借助微信公众平台（即微信公众号）深入开展阅读推广已成为工作中的重要部分。

如何将微信公众平台更好地运用于图书馆工作中，并深入有效地推进阅读推广是当前图书馆需要深入思考的问题。学者们从理论与实践的角度对图书馆开展的阅读推广活动进行了研究，其中也包含了使用微信公众平台开展阅读推广的相关研究，多数学者侧重于对图书馆微信公众平台的开通情况、菜单设置内容等进行研究，鲜有学者从文本内容角度针对图书馆发布的阅读推广文章进行深入细致的挖掘分析。因此本文选取我国31家省级公共图书馆，将其微信公众平台发布的关于阅读推广的文章及相关数据进行采集和统计，运用文本挖掘方法对阅读推广的内容主题及其演化特征进行挖掘分析。

2 微信阅读推广内容界定

当前已有不少学者给出阅读推广的定义。例如范并思认为图书馆阅读推广是我国图书馆界对图书馆营销推广和图书馆新型阅读服务的统称[2]。张怀涛将阅读推广定义为：阅读推广顾名思义就是推广阅读，简言之就是社会组织或个人为促进人们阅读而开展的相关活动，也就是将有益于个人和社会的阅读活动推而广之；详言之就是社会组织或个人，为促进阅读这一人类独有的活动，采用相应的途径和方式，扩展阅读的作用范围，增强阅读的影响力，使人们更有意愿、更有条件参与阅读的人文活动和事业[3]。王波将阅读推广表述为：阅读推广就是为了推动人人阅读，以提高人类文化素质、提升民族软实力、加快各国富强和民族振兴的进程为战略目标，而由各国的机构和个人开展旨在培养民众的阅读兴趣、阅读习惯，提高民众的阅读质量、阅读能力、阅读效果的活动[4]。综上可知，阅读推广的内

* 本文系国家社会科学基金一般项目"中文学术领域命名实体的知识图谱构建研究"（项目编号：18BTQ072）和山西省高等学校哲学社会科学研究项目"'互联网+'背景下山西省中小型工业企业营销模式创新研究"（项目编号：2016204）研究成果之一。

作者简介：李倩，副教授，博士，硕士生导师，中国科学技术信息研究所在职博士后，E-mail：86873657@qq.com；王帅，硕士研究生。

收稿日期：2021-08-23 修回日期：2021-12-22 本文起止页码：72-83 本文责任编辑：王传清

涵较为宽泛,它不仅指可直接或间接阅读的书籍资源的推广,凡是有助于培养读者阅读方法、提升阅读素养、激发阅读兴趣、养成阅读习惯以及带给读者阅读享受等的相关信息均可认为是阅读推广内容。

图书馆借助微信公众平台所发布的内容并非都是阅读推广内容,其中还包括图书馆新闻公告、闭馆开馆工作通知、地区新闻等内容,因此需要对阅读推广内容进行界定和识别。在现有的相关研究中已有学者对图书馆微信公众平台中的阅读推广工作的内容进行识别和分类。例如万慕晨等指出阅读推广不应只局限于图书推荐、读书征文比赛、名家讲座等活动,还应扩大其范围,如将书签设计、摄影展览、有奖竞答等视为阅读推广的活动,因此认为图书馆微信公众平台中的阅读推广信息除了图书馆新闻公告、相关工作通知、信息查询与服务及其他较难界定的信息外,都属于与阅读推广相关的内容范围[5]。蔡丽萍等将图书馆微信公众平台中的阅读推广内容界定为除了图书馆新闻公告、相关工作通知、会议报道、信息查询与服务说明、节日祝福及其他较难界定的信息外,其余均列为阅读推广内容范畴[6]。王宝英等认为阅读推广工作包括动态活动推广和静态内容推广两方面,利用微信发布的阅读推广动态活动预告仅起到广告宣传作用,实质性的阅读推广功效却不大,因而将对高校图书馆微信公众平台发布的阅读推广内容探讨重点放在图书推荐、借阅排行榜、各类书单、书评、新书推荐、期刊推荐、电子图书推荐书单、美文或心灵鸡汤、动态活动成果展示、读书趣事或方法、知识推介等静态阅读推广资源上[7]。

结合阅读推广定义及上述学者观点,本文中所研究的图书馆微信公众平台阅读推广内容可界定为图书馆除了可直接或间接阅读的书籍资源的推广外,凡是有助于培养读者的阅读方法、提升阅读素养、激发阅读兴趣、养成阅读习惯以及带给读者阅读享受等的相关信息均可认为是阅读推广内容。界定和识别阅读推广信息可依据两个方面进行:①读者可直接阅读(微阅读、书、报刊)。②信息能够促进用户阅读素养(兴趣、习惯、方法、能力、道德)的提升。

3 研究设计

3.1 研究对象

本文选取国内 31 个省级[8]公共图书馆的微信公众平台为研究对象,采集工作开始于 2020 年 8 月 25 日,数据的时间段选取 2019 年 11 月 1 日至 2020 年 7 月 31 日这 9 个月共计 273 天,借助清博智能平台及微信手机客户端对数据进行调查采集。清博智能数据库中包含了微信公众平台中的大部分数据并提供了数据监测功能。31 个省级公共图书馆微信公众号名称和开通日期如表 1 所示:

表 1 省级公共图书馆微信公众号基本情况(依开通日期排序)

序号	图书馆名称	开通日期	序号	图书馆名称	开通日期	序号	图书馆名称	开通日期
1	新疆图书馆	2013－05－07	12	西藏自治区图书馆	2014－04－18	23	贵州省图书馆	2014－10－24
2	湖北省图书馆	2013－05－25	13	重庆图书馆	2014－04－22	24	浙江图书馆	2014－10－28
3	首都图书馆	2013－11－23	14	福建省图书馆	2014－04－23	25	天津图书馆	2015－03－04
4	上海图书馆	2013－12－20	15	南京图书馆	2014－05－20	26	江西省图书馆	2015－04－20
5	陕西省图书馆	2014－01－26	16	广西壮族自治区图书馆	2014－05－28	27	青海省图书馆	2015－04－23
6	黑龙江省图书馆	2014－02－12	17	辽宁省图书馆	2014－06－09	28	云南省图书馆	2015－04－28
7	湖南图书馆	2014－02－15	18	河南省图书馆	2014－06－10	39	安徽省图书馆	2015－07－09
8	山东省图书馆	2014－03－13	19	内蒙古自治区图书馆	2014－07－10	30	四川省图书馆	2015－07－16
9	广东省立中山图书馆	2014－03－18	20	河北省图书馆	2014－08－07	31	宁夏图书馆	2016－03－15
10	海南省图书馆	2014－03－28	21	吉林省图书馆	2014－09－17			
11	山西省图书馆	2014－03－31	22	甘肃省图书馆	2014－10－15			

3.2 研究方法

本文采用 LDA(Latent Dirichlet Allocation)主题模型进行阅读推广内容主题特征分析。LDA 主题模型是一种非监督机器学习技术,利用三层贝叶斯概率模型识别大规模文档中潜藏的主题信息[9]。它基于如下假设:①文档集合中存在 K 个主题且主题之间相互独立;②每个文档由 K 个主题随机混合组成,且主题参数服从 Dirichlet 分布;③每个主题是特征词上的多项分布,该多项分布的参数服从 Dirichlet 分布。它的主要思想是:一篇文档以一定的概率选择了某个主题,并从该主题中以一定概率选择某个词语而构成,即一篇文档代表若干主题构成的一个概率分布,而每一主题又代表若干词语构成的一个概率分布。LDA 的模型计算结果可以得到文档－主题和主题－词语 2 个概率分布。文

档－主题概率分布获得每个主题下的文档支持权重，权重越大表示该文档与主题的关联度越大；主题－词语概率分布由一系列特征词及其在该主题中出现的概率值表示，由此反映每个主题的内部结构，特征词概率值越大，该词对该主题的贡献度越高[9-10]。

在 LDA 模型中，一篇文档的“文档－词项”的生成过程如图 1 所示，首先从 Dirichlet 分布 α 中选择生成文档的主题分布 θ_m，也就是说，主题分布 θ_m 是由超参数为 α 的 Dirichlet 分布生成，然后从 θ_m 取样生成文档的第 n 个词的主题 $z_{m,n}$，接着从 Dirichlet 分布 β 取样生成主题 $z_{m,n}$ 对应的多项词语分布 φ_k，最后从 φ_k 中采样生成单词 $W_{m,n}$。将上述过程重复 n 次，便可生成一篇文档的“文档－词项”[10]。

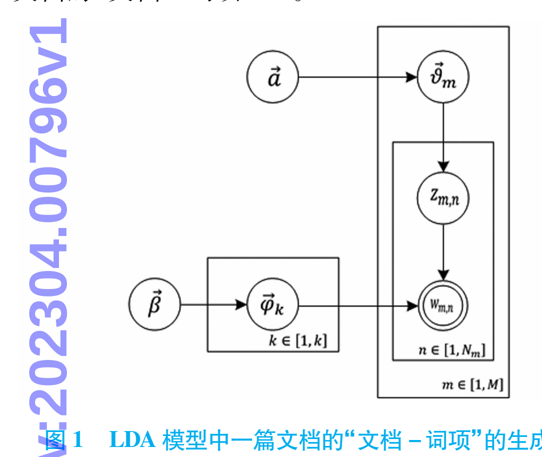


图 1 LDA 模型中一篇文档的“文档－词项”的生成过程

在 LDA 主题模型的实验过程中，通常需要设定划分主题个数。D. M. Blei 最先采用困惑度（Perplexity）确定最佳主题数目 K[11]，在信息论中，困惑度是用来度量一个概率分布或概率预测样本的好坏程度。困惑度已成为判断最优主题数的重要指标。计算公式如（1）（2）所示：

$$Perplexity(D) = \exp\left\{ - \frac{\sum_{d=1}^{M} \log p(w_d)}{\sum_{d=1}^{M} N_d} \right\} \quad \text{公式(1)}$$

$$p(w_d) = \prod_{i=1}^{(N_d)} \sum_z p(w_{d,i}|z)\ p(z|d) \quad \text{公式(2)}$$

其中，N_d 为词汇集合，也代表文档的长度，M 为文档数，$p(w_d)$ 为文档生成概率，$w_{d,i}$ 为文档 d 中的第 i 个词，z 为特定主题。结合公式（1）中的“exp( )”为以自然常数 e 为底的指数函数式子、邱均平和沈超的论文[12]、网友讨论及 python 代码中“log( )”形式的函数默认以 e 为底[13]，认为公式（1）中的 log 的底为 e。公式（2）中 $\sum_z p(w_{d,i}|z)\ p(z|d)$ 为文档 d 中第 i 个词汇的生成概率，由于是词袋模型，每个词汇的概率条件独立，文档 d 的生成概率 $p(w_d)$ 为其文档中每个词汇产生概率之积。

LDA 主题模型是文本主题分类中常使用的一种分类模型，目前已有多个软件可实现文本数据分析中的主题建模和挖掘，本文主要借助 Python 中 Scikit-learn 第三方工具中的 LatentDirichletAllocation 库实现建模，最终得到阅读推广内容主题－特征词分布及文档－主题分布，从而实现阅读推广内容的主题分析。数据采集及数据分析流程如图 2 所示：

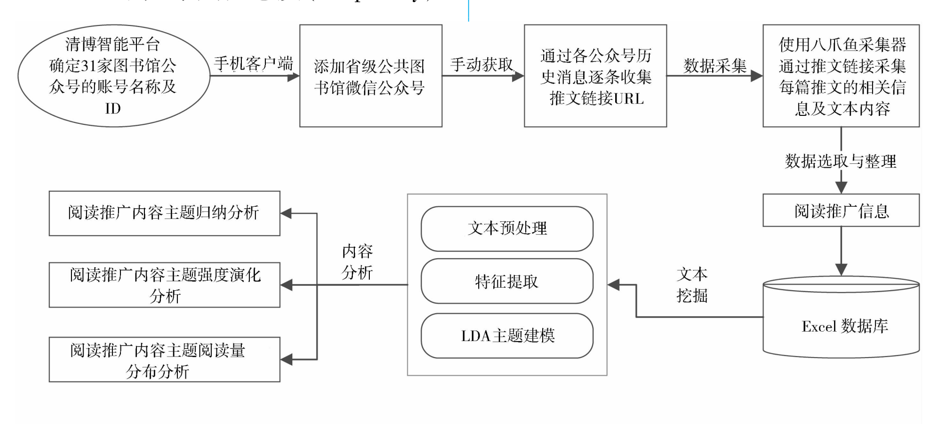


图 2 图书馆微信公众号数据采集及分析流程

3.3 数据采集与选取

由于微信版本的不断更新以及微信网页版无法登陆等问题对微信数据的采集带来一定的困难，且目前也尚未开发出支持微信公众号推送文章及相关数据的采集工具软件，若通过手机软件抓包或自主构建一套爬虫系统技术门槛较高，较难实现，因此，本文通过手动和采集工具相结合的方式进行数据采集。采集步骤为：①使用微信手机客户端添加 31 家省级公共图书馆的微信公众号。通过微信手机客户端中的“添加朋友”功能区逐一搜索 31 家省级公共图书馆名称，为了

保证微信公众号的真实可靠性,避免所关注公众号为虚假账号,笔者通过清博智能平台核查账号主体注册信息是否为该省公共图书馆官方账号,然后对其进行关注。②使用微信手机客户端手动获取公众号推送文章的 URL 链接、阅读数和在看数。在微信公众号的“历史消息”列表中,逐一点击并打开推送文章页面,首先点击推文右上角隐藏菜单栏中的“复制链接”图标,复制当前推文的链接并保存至 Excel 中;其次滑到文章底部获取此推文的阅读数和在看数。将同一公众号内推文的链接、阅读数和在看数保存至一个 Excel 文件中,共获得 31 个文件,11 131 条推文记录。③使用八爪鱼采集器获取推文文本内容及相关信息。将获取的推文 URL 导入八爪鱼采集器中,通过设置自定义采集信息模块及采集流程批量获取推文相关信息,其中包括推文标题、发布时间和正文内容。采集完成后将所有采集的数据使用 Excel 进行整合,每个省级公共图书馆的数据是一个单独的文件,采集的湖南图书馆微信公众号部分数据如图 3 所示:

1		标题	日期	正文
2	湖南省图书馆	“张宗昌开始人传人了”	7月31日	“张宗昌开始人传人了” 微信号hntsg1904功能介绍读者服务、活动预告、
3	湖南省图书馆	“小妈妈”们，何以为家？	7月31日	“小妈妈”们，何以为家？ 微信号hntsg1904功能介绍读者服务、活动预告
4	湖南省图书馆	八月公益文化惠民活动早知道！	7月31日	八月公益文化惠民活动早知道！ 微信号hntsg1904功能介绍读者服务、
5	湖南省图书馆	为什么还有人在淘宝买老人机	7月30日	为什么还有人在淘宝买老人机 微信号hntsg1904功能介绍读者服务、活动
6	湖南省图书馆	动物界唯一的活菩萨，而且据说还很好吃	7月30日	动物界唯一的活菩萨，而且据说还很好吃 微信号hntsg1904功能介绍读者
7	湖南省图书馆	【赠书】后怕！这些藏在孩子身边的危险，爸妈一定要注意！	7月30日	【赠书】后怕！这些藏在孩子身边的危险，爸妈一定要注意！ 微信号
8	湖南省图书馆	《三十而已》爆火：比起出轨离婚，我更怕活成王漫妮	7月29日	《三十而已》爆火：比起出轨离婚，我更怕活成王漫妮 微信号hntsg1904
9	湖南省图书馆	48位高考状元身份揭秘：每个横空出世的奇迹，都是父母在奋力托举	7月29日	48位高考状元身份揭秘：每个横空出世的奇迹，都是父母在奋力托举 微
10	湖南省图书馆	五六岁小男孩被妈妈送到派出所：我来“自首”了……	7月28日	五六岁小男孩被妈妈送到派出所：我来“自首”了…… 微信号hntsg1904功
11	湖南省图书馆	2020年高考状元竟然是这位神仙小姐姐！总分725，语文146，数学149，十年难遇！	7月28日	2020年高考状元竟然是这位神仙小姐姐！总分725，语文146，数学149，
12	湖南省图书馆	你都不敢出发，凭什么嘲笑骑行西藏的人	7月27日	你都不敢出发，凭什么嘲笑骑行西藏的人 微信号hntsg1904功能介绍读者
13	湖南省图书馆	一张成绩单，把家长分为三六九等，你是哪一等？	7月27日	一张成绩单，把家长分为三六九等，你是哪一等？ 微信号hntsg1904功能
14	湖南省图书馆	这个女人的朋友圈，我不敢看	7月26日	这个女人的朋友圈，我不敢看 微信号hntsg1904功能介绍读者服务、活动
15	湖南省图书馆	史上最难高考出分了：请抱抱那个考砸了的孩子，他们是来报恩的	7月26日	史上最难高考出分了：请抱抱那个考砸了的孩子，他们是来报恩的我是
16	湖南省图书馆	周末读诗丨李清照这8首诗词，说尽一生，美到极致！	7月26日	周末读诗丨李清照这8首诗词，说尽一生，美到极致！ 微信号hntsg1904
17	湖南省图书馆	“河南李子柒”毕业为父变“网红”，192w粉丝坐等更新，连央视都点赞…	7月25日	“河南李子柒”毕业为父变“网红”，192w粉丝坐等更新，连央视都点赞…
18	湖南省图书馆	中国第一个驻美总领事馆，为什么是休斯敦？	7月25日	中国第一个驻美总领事馆，为什么是休斯敦？ 微信号hntsg1904功能介绍
19	湖南省图书馆	赠书|《乘风破浪的姐姐》爆火：会吃的女人不会老	7月25日	赠书|《乘风破浪的姐姐》爆火：会吃的女人不会老 微信号hntsg1904功能
20	湖南省图书馆	62岁李宗盛孤独庆生：“我终于放下了你，你却留我独自白头”	7月24日	62岁李宗盛孤独庆生：“我终于放下了你，你却留我独自白头” 微信号

图 3　湖南图书馆微信公众号数据示例

在信息采集过程中存在数据无法采集成功的现象,这是由于微信公众号中推文发布的形式不仅仅只是文字,还包括图片、音频、视频形式的信息,除此之外,微信公众号中文字的排版格式在一定程度上也会影响信息采集的效果。因此,数据选取的第一步就是去除正文内容采集数据为空的信息,共计 246 条。图书馆借助微信公众平台所发布的信息并非都是阅读推广信息,其中还包括新闻公告、图书馆工作通知等内容,因此需要对阅读推广信息进行筛选。笔者采用人工方式,按照上文中对阅读推广信息的概念界定进行逐条筛选,通过阅读文章的标题进行判断,最终筛选出非阅读推广信息 1 009 条,阅读推广信息数据共计9 876 条。

3.4　阅读推广内容的 LDA 主题发现

3.4.1　文本预处理

(1)去除特殊符号及格式。在采集的原始文本数据中通常包含英文、数字、特殊符号和表情等对文本分析贡献不大且对机器的理解存在干扰的成分,因此,笔者采用 Excel 中的智能工具箱,批量去除文本中的数字、英文、符号,形成无符号的纯文本形式。

(2)构建停用词表。在使用机器学习文本分类算法的过程中,去除停用词是非常重要且关键的部分,如果包含太多杂词,会导致算法效果不尽人意。在进行 LDA 模型的构建中,需要从文本中提取文档的特征词,但在非结构化的文档集中常常存在频率较高却无实际含义的词汇,比如介词、连词等在句子中起到连贯作用的词语,还有一些例如“他的”“我的”“他们”等指示代词不能够代表文本实际意义成为特征词,通常将这些词在文本处理时称为停用词。目前已经有专门的停用词表,但在实际操作中,针对不同的文档集合直接使用已有的停用词表并不能达到较好的分类效果,因此,需要针对文档集人工构建停用词表。本文在“百度停用词表”的基础上,结合文档词性标注及词频统计,将出现频次高的词语,例如“图书馆”“阅读”“读书”等删去,同时将连词、叹词、数词、介词、代词等词也纳入停用词表,最终构建的停用词表包含 2 318 个词。

3.4.2　特征提取

文本特征提取是 LDA 建模过程中必不可少的环节,计算机本身并不能够读懂文字含义,因此在文本处理或文本挖掘过程中需要通过计算文字权重进而转换成可以量化的特征词,方便进一步的数据挖掘。本文使用 TF-IDF(Term Frequency Inverse Document Frequency)进行文本特征提取。TF-IDF 是一种信息检索的加权技术,其中 TF 代表词频,IDF 代表逆向文档频率,它是以特征词在文档中出现的次数除以包含该特征词的文档数作为该词的权重,常用来做文本分类。使用 Py-

thon 中的 Scikit-learn 模块所含的 TfidfVectorizer 参数进行 TF-IDF 特征提取，由于需要处理的文本数据量较大，其中包含有大量词汇，很多词汇对于主题抽取意义不大且在文本处理中耗时较长，因此，设定 n_features = 3 000，从文本中提取 3 000 个最重要的特征关键词，通过 print(tf)输出结果得到了 9876 ×1136 的特征词稀疏矩阵。

3.4.3　LDA 主题建模

在停用词表及主题特征提取的基础上采用 Scikit-learn 包中 LatentDirichletAllocation 库构建主题模型，其中超参数 α 和 η 选取默认值，最大迭代次数 max_iter =1 000。LDA 主题模型需要预先设定主题数目，较权威的主题数确定方法是采用困惑度进行计算，困惑度是判断最优主题数的重要指标，计算使用 lda. perplexity 输出最佳主题数。由图 4 可知，主题数为 9 时，模型的困惑度较低，同时结合 pyLDAvis 可视化结果(见图 5 左边区域)发现，当主题数 K =9 时，主题之间的重叠较少，分类效果较好，因此，最终将主题数 K 确定为 9。图 5 的右边区域展示与 Topic3 相关性较大的前 30 个词汇。可以在图 5 左上角的“Selected Topic”文本框中输入不同的序号以展示不同主题下相关性较大的排名前 30 的词汇。

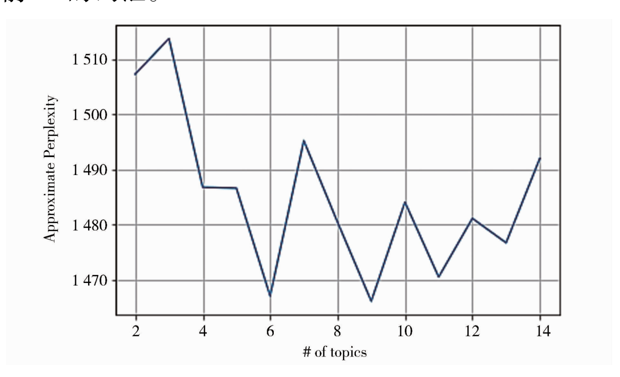


图 4　困惑度计算结果

注：横坐标表示主题个数，单位为“个”；纵坐标表示困惑度大小，没有具体的单位，其大小是判断取舍主题数的依据，取困惑度值最小的时候对应的主题数

图 5　pyLDA 结果可视化

4　数据结果与分析

4.1　阅读推广内容主题归纳分析

通过 LDA 主题分类得到图书馆微信阅读推广内容的 9 个主题以及每个主题的词项分布，将每个主题下排名前 15 的高概率特征词进行整理，并依据每个主题下的高概率特征词进行场景描述，总结归纳出最符合当前主题下高概率特征词的主题标识。例如在 Top-

ic0 中,电子书、共享、检索、疫情、网站等高概率特征词贴合疫情期间全民宅在家,利用网络在线推荐、获取相关数字资源服务的情境,故将 Topic0 定义为"数字资源推广与获取"。在 Topic1 中,少儿、小朋友、亲子、家长、共读等这些高概率特征词与家庭学习较为密切,更好地体现了亲子学习教育这类场景,因此将 Topic1 定义为家庭与亲子教育。在 Topic4 中,诗词、文化遗产、民俗、风俗等这类词符合对中国传统诗词文化、民俗节日的描述,因此将 Topic4 定义为传统文化。依次类推,最终将省级公共图书馆阅读推广主题归纳结果分为九大类,如表 2 所示:

表 2　主题 – 高概率特征词分布

编号	主题标识类别	主题的前 15 个高概率特征词
Topic0	数字资源推广与获取	古籍 全民 电子书 书法 疫情 奖项 上线 共享 检索 专题 海量 少儿 市民 网站 名家
Topic1	家庭与亲子教育	少儿 小朋友 亲子 家长 共读 少儿馆 教育 培养 成长 姐姐 兴趣 英文 童年 讲故事 青少年
Topic2	生活、情感	父亲 成长 儿子 女性 爱情 年轻人 思考 动物 女儿 书单 青年 旅行 长大 情绪 孤独
Topic3	科学普及	少儿 挑战 动画 健康 小学 全民 编排 问答 链接 小伙伴 宅家 抽奖 科普 疫情 宇宙
Topic4	传统文化	诗词 博物馆 文化遗产 民俗 民间 习俗 考古 文物 工艺 美食 节日 古代 时节 植物 风俗
Topic5	疫情防护与健康	疫情 防控 肺炎 冠状病毒 病毒 抗疫 口罩 防护 感染 健康 加油 人民 一线 致敬 预防
Topic6	诗歌艺术鉴赏	音乐 诗歌 诗人 国学 诗词 名师 法国 古典 美学 古代 喜剧 风格 大师 艺术家 表演
Topic7	爱国教育主题	中国共产党 人民 初心 庆祝 同志 和平 爱国 青年 使命 胜利 纪念 战争 社会主义 奋斗 信仰
Topic8	经济、政治、科技、哲学	建设 教育 经济 理论 专家 习近平 政治 领域 导师 哲学 法律 总书记 社会主义 制度 科技

由表 2 可知,Topic0 为数字资源推广与获取,图书馆积极抓住新媒体发展趋势,积极利用互联网与社交媒体带来的交互功能和传递信息快速的特点,加强数字资源的宣传,积极推动数字资源服务。Topic1 为家庭与亲子教育,公共图书馆利用其资源和环境优势开展家庭教育,引导家长参与到孩子的阅读中去,有助于亲子关系的养成。Topic2 为生活、情感,其推广对象主要关注青年一代,主题内容与青年生活、情感、奋斗、励志息息相关,且推广内容大多能够引起青年当下的共鸣。Topic3 为科学普及,这一主题主要以健康和科学为科普对象,阅读推广活动主要通过答题互动、讲座的形式来开展。Topic4、Topic6、Topic7 和 Topic8 表明省级公共图书馆通过阅读推广活动积极向社会传承中国优秀传统文化和国学经典,加强爱国主义教育、社会主义制度、经济、政治、哲学、科学、法律等内容推广。Topic5 为疫情防护与健康,2020 年初突发重大公共卫生安全事件,图书馆积极承担并履行社会责任,通过线上开展的阅读推广活动,向社会提供数字资源服务和在线讲座活动。

4.2　阅读推广内容主题强度及演化分析

4.2.1　阅读推广内容主题强度分析

阅读推广内容主题强度分布情况如表 3 所示,$\theta_{i,j}$ 中 i 表示阅读推广文章的序号,也可将其看作是一条阅读推广信息,j 表示其所属的 Topic。其中,第 1 条、第 2 条、第 3 条、第 9 875 条阅读推广信息在 Topic2 中得到了概率分布的最大值,第 9 874 条阅读推广信息在 Topic8 中得到的概率最大,第 9 876 条阅读推广信息在 Topic6 中得到的概率最大。

表 3　文档 – 主题分布

$\theta_{i,j}$	Topic0	Topic1	Topic2	Topic3	Topic4	Topic5	Topic6	Topic7	Topic8
1	0.023 1	0.023 0	**0.815 6**	0.023 1	0.023 0	0.023 0	0.023 0	0.023 1	0.023 0
2	0.012 9	0.012 9	**0.896 5**	0.012 9	0.012 9	0.012 9	0.012 9	0.012 9	0.012 9
3	0.020 1	0.020 1	**0.838 9**	0.020 1	0.020 1	0.020 1	0.020 1	0.020 1	0.020 1
…									
9 874	0.017 6	0.017 5	0.030 9	0.017 5	0.017 5	0.017 6	0.017 5	0.017 5	**0.846 2**
9 875	0.023 6	0.023 6	**0.811 0**	0.023 6	0.023 7	0.023 6	0.023 6	0.023 6	0.023 6
9 876	0.027 9	0.027 8	0.060 1	0.028 2	0.027 9	0.027 8	**0.744 7**	0.027 8	0.027 8

由表 3 可知,第 1 条推文的标题为《浙图锦鲤 2.0 全面升级!海量书籍、Kindle 阅读神器、旅行套餐……带上诗书去远行》[14],主要说明"浙图锦鲤 2.0"的书籍推荐活动与"文旅融合"的旅游路线推介活动,其主题与表 2 中的 Topic2 的"书单""旅行"等关键词相契合。第 2 条推文标题为《如果不是因为一场意外,大唐诗坛一哥或许不是李白》[15],主要是对王勃的短暂而辉煌一生及其所写的诗与文章进行介绍,其主题与表 2 中

的 Topic2 中的"成长""青年""长大""孤独"等关键词相契合。第 3 条推文标题为《中国南海首次发现鲸落：一鲸落，万物生！这是世界上最浪漫的死亡……》[16]，介绍一头鲸在死亡后的长达一百年左右的时间，养活 1 万个生物体，形成繁荣庞大的生物群落。在启迪读者智慧的同时激发读者对"鲸落"相关内容的阅读兴趣，其主题与表 2 中的 Topic2 的"动物"等关键词相契合。第 9 875 条推文标题为《大才子造了一个谣，骗了我们几百年》[17]，主要探讨张飞长相是粗犷还是俊美，其女儿是否貌美，并以《立马铭》《蜀中名胜记》《刁斗铭》《骄儿诗》《三国志》等为探讨考据的依据，其主题与表 2 中 Topic2 的"父亲""女儿"等关键词相契合。因此，以上阅读推广信息较大概率内容是关于 Topic2 的。第 9 874 条推文标题为《四讲四爱，激发西藏脱贫攻坚动能》[18]，介绍了"四讲四爱"群众教育实践活动，激发读者对科技文化和脱贫攻坚时期的政治文献的阅读兴趣，其主题与表 2 中的 Topic8 的"教育""习近平""政治""总书记""科技"等关键词相契合，因此该条信息较大概率属于 Topic8。第 9 876 条推文为《一"荐"倾心 | <读懂古典诗词>》[19] 主要进行诗词的鉴赏与推介，与表 2 中 Topic6 的"诗歌""诗人""诗词""古典""国学"等关键词相契合。因此，第 9 876 条阅读推广信息较大概率内容是关于 Topic6 的。值得注意的是，第 3 条推文未归入到 Topic3"科学普及"，主要是因为文章以诗词的艺术形式进行展现，并且未明确表示是科普类文章，科普类相关关键词偏少。

在文档 – 主题概率分布的基础上进行统计得到如图 6 所示的各个主题类别下阅读推广信息数量（即篇数）。

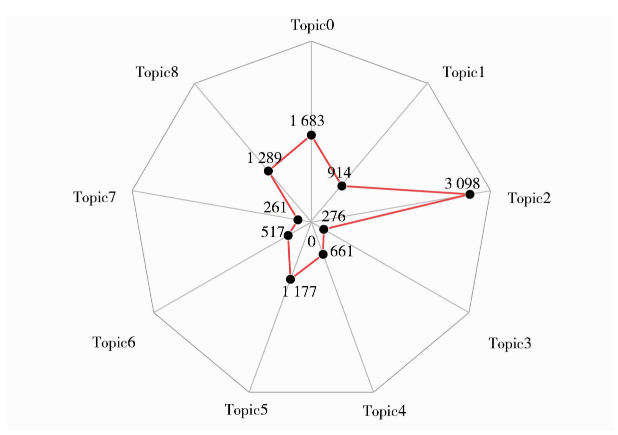


图 6　各个主题类别对应的阅读推广信息数量（单位：篇）

由图 6 可知，各个主题类别对应的阅读推广信息发布数量从多到少依次为：Topic2（3 098 条）、Topic0（1 683 条）、Topic8（1 289 条）、Topic5（1 177 条）、Topic1（914 条）、Topic4（661 条）、Topic6（517 条）、Topic3（276 条）、Topic7（261 条）。

4.2.2　阅读推广内容主题强度演化分析

阅读推广主题强度演化即阅读推广主题强度随着时间的推移所显示出的阅读推广发布信息数量（也简称为阅读推广数）变化。阅读推广的九大主题从 2019 年 11 月 – 2020 年 7 月时间段内演化趋势见图 7。

从整体演化情况来看，每个主题的阅读推广数均呈现出不同的波动情况，Topic5（疫情防护与健康）的阅读推广数受疫情影响，在 2020 年 1 月开始呈明显的上升趋势，1 月到 2 月时间段内数量呈现迅猛增长，2 月份达到发布数量的月峰值 434 条；同样受疫情影响，Topic1（家庭与亲子教育）、Topic2（生活、情感）、Topic4（传统文化）、Topic6（诗歌艺术鉴赏）、Topic7（爱国教育主题）和 Topic8（经济、政治、科技、哲学）等主题各自的阅读推广数在 2 月份均呈现明显的下降趋势。Topic0（数字资源推广与获取）和 Topic3（科学普及）两个主题的阅读推广数未受到影响反而呈现一定的上升趋势。

伴随着疫情防控措施的落实及防疫成果显著，除 Topic5 外各主题的阅读推广数在 4 月份逐渐回升并达到小高峰，Topic5（疫情防护与健康）的阅读推广数则呈明显加速下降趋势。各主题的阅读推广数在 4 月份回升并达到小高峰，一方面是因为疫情得到有效控制；另一方面是由于 4 月 23 日为世界读书日，各图书馆都将 4 月份作为读书月并设计和组织多种形式的以阅读为主题的读书活动。如南京图书馆开展了"4.23 云共读 | 不学诗，无以言：书香战'疫'润心灵"的主题活动、四大名著争锋在线知识竞赛等活动等。Topic4（传统文化）的另一个小高峰出现于 1 月份，最大的高峰出现于 6 月份，与传统佳节元旦、春节、端午节的来临关系比较大。

Topic7（爱国教育主题）的阅读推广数受疫情影响波动幅度较小，可忽略不计，从 2019 年 11 月 – 2020 年 6 月较为平稳，7 月呈明显的上升趋势，图书馆将 7 月作为红色文化月，通过推荐红色阅读文献、开展红色经典朗诵比赛、地方特色历史介绍、革命文化宣传等线上阅读推广活动向大众传递爱国主义精神。

4.3　阅读推广内容主题阅读量分布分析

为了分析每个阅读推广主题的读者阅读情况，以该主题下的所有阅读推广信息的阅读量除以阅读推广

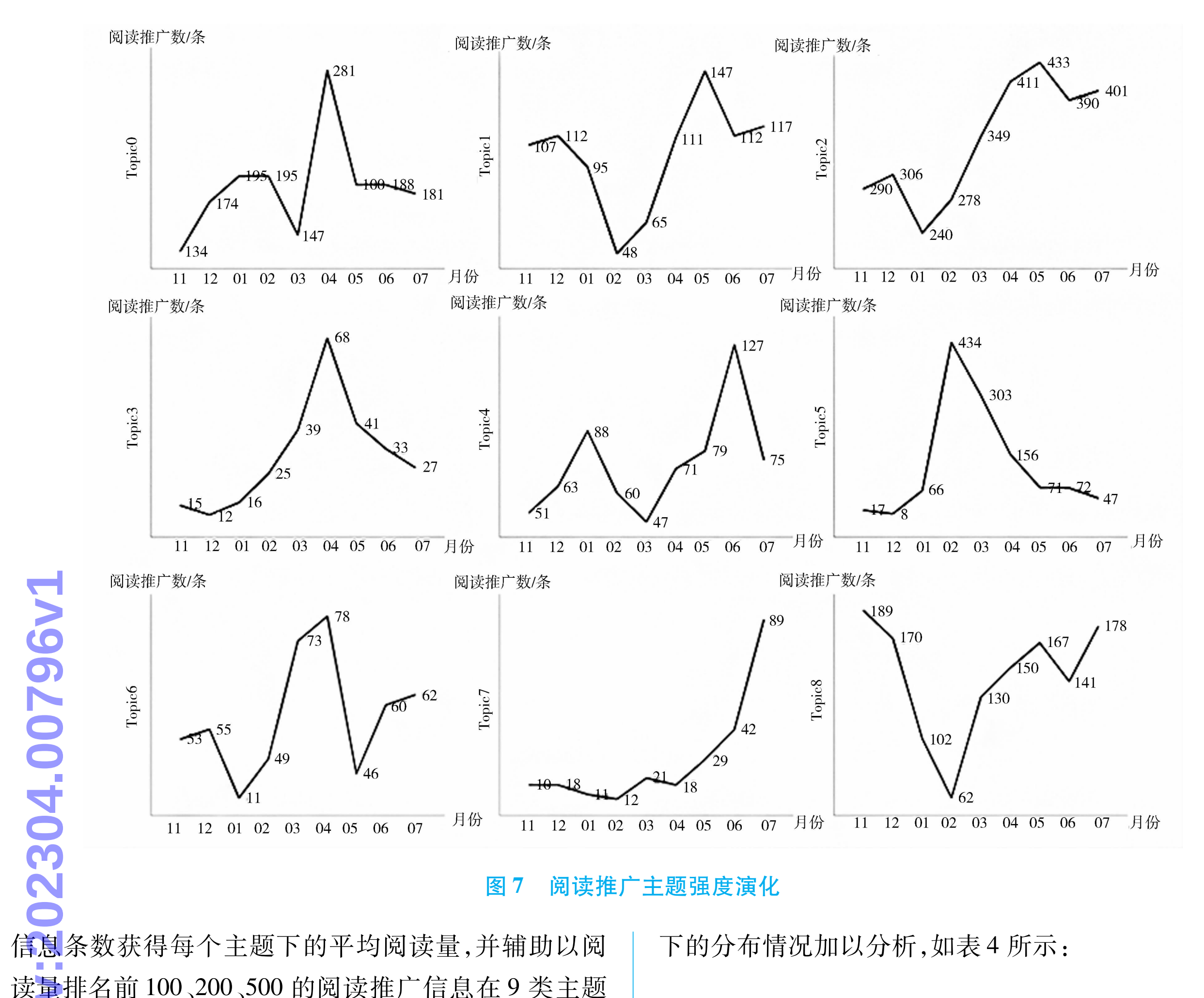


图 7　阅读推广主题强度演化

信息条数获得每个主题下的平均阅读量，并辅助以阅读量排名前 100、200、500 的阅读推广信息在 9 类主题下的分布情况加以分析，如表 4 所示：

表 4　主题－阅读量分布

Topic 序号	平均阅读量(阅读量/次)	阅读量排名前 100 的阅读推广信息分布情况/次	阅读量排名前 200 的阅读推广信息分布情况/次	阅读量排名前 500 的阅读推广信息分布情况/次	阅读量排名前 1 000 的阅读推广信息分布情况/次
Topic0	953.923 945 3	4	7	29	84
Topic1	1 008.850 109	8	15	18	33
Topic2	2 693.974 822	76	145	342	637
Topic3	666.442 029	0	0	4	7
Topic4	774.075 643	1	2	5	22
Topic5	1 452.297 114	3	9	61	140
Topic6	633.621 359 2	1	2	5	12
Topic7	376.605 364	0	0	0	2
Topic8	1 017.044 22	7	20	36	63
依据平均阅读量大小对各主题排序(Topic 序号)			2、5、8、1、0、4、3、6、7		
依据阅读量排名前 100 的阅读推广信息分布数量大小对各主题排序(Topic 序号)			2、1、8、0、5、4 与 6 并列、3 与 7 并列		
依据阅读量排名前 200 的阅读推广信息分布数量大小对各主题排序(Topic 序号)			2、8、1、5、0、4 与 6 并列、3 与 7 并列		
依据阅读量排名前 500 的阅读推广信息分布数量大小对各主题排序(Topic 序号)			2、5、8、0、1、4 与 6 并列、3、7		
依据阅读量排名前 1 000 的阅读推广信息分布数量大小对各主题排序(Topic 序号)			2、5、0、8、1、4、6、3、7		

由表 4 可知，Topic2(生活、情感)在几项指标中均排名第 1，这说明 Topic2(生活、情感)总体受欢迎。Topic5(疫情防护与健康)在阅读量排名前 100 的信息中拥有的信息分布数量排名第 5，而在阅读量排名前 1 000的信息中信息分布数量排名上升到第 2，其平均阅读量大小总体排名也是第 2。这说明 Topic5(疫情

防护与健康)总体上受欢迎,但是在前100名和前200名的最热排序中名次不拔尖。Topic8在阅读量排名前100的信息中拥有的信息分布数量排名第3,中间曾一度跃升至第2名,但是在阅读量排名前1 000的信息中的信息分布数量排名下降到第4,其平均阅读量大小总体排名是第3。这说明Topic8(经济、政治、科技、哲学)总体受欢迎程度排在第3名,虽有波动,但较为稳定。Topic1(家庭与亲子教育)在阅读量排名前100的信息中拥有的信息分布数量排名第2,而在阅读量排名前1 000的信息中的信息分布数量排名下降到第5,其平均阅读量大小总体排名是第4。这说明Topic1(家庭与亲子教育)有部分信息热度很高,但是总体比较受欢迎。Topic0(数字资源推广与获取)总体排名虽有变化,但总体保持在九类主题的中位水平。Topic4(传统文化)、Topic3(科学普及)、Topic6(诗歌艺术鉴赏)、Topic7(爱国教育主题)总体阅读量排名靠后。

5 研究结论

5.1 阅读推广内容主题归类体现出的特征

5.1.1 推广内容涉及范围广,彰显服务于大众的理念

目前省级公共图书馆在阅读推广的内容主题方面较为广泛,涵盖了生活、情感、文化、艺术、健康、政治、经济、科技、哲学等多个领域,涉及大众日常的学习、工作与生活的多个方面。体现出省级公共图书馆服务于大众的理念。从推广对象来看,有时以儿童为主,有时以青少年为主,有时则面向全体大众,可以说阅读推广对象类型涉及较多。与此同时,图书馆根据不同的目标人群在内容上具有一定的主题特征,例如Topic0(数字资源推广与获取)面向的推广人群为全体民众,而Topic1(家庭与亲子教育)面向的推广人群主要为幼儿与青少年,Topic2(生活、情感)面向的推广人群为青年或女性等群体。

5.1.2 注重传播文化经典,弘扬优秀文化

图书馆通过开展优秀传统文化的阅读推广活动,能够帮助青少年了解和学习经典国学文化,从优秀的传统文化中汲取养分。图书馆推广的阅读内容有《诗经》《孟子》《论语》等经典书籍。通过侧重图画阅读的儿童绘本引导少年儿童进行阅读理解;通过诵读经典、诗歌鉴赏的活动方式吸引并组织带领少年儿童吟诵和理解中国传统文化之美;通过开展线上+线下的文化大讲堂活动,传授经典文化知识,营造浓厚的学习氛围,让经典国学走向大众,充分发挥了领读中华传统文化经典的带头作用。

5.1.3 紧扣时代发展,积极承担社会责任

省级公共图书馆阅读推广内容不局限于传承中华优秀传统经典,同时能够紧扣时代发展趋势,一是体现在弘扬社会主义经济、制度和法律层面,各省级图书馆通过讲座普及法律、对党的重要会议精神解读等方式向大众传递核心价值观;二是体现在疫情期间,图书馆作为培育文化自信的重要场所,积极承担社会责任,通过开展线上讲座、线上培训、线上主题阅读、在线展览等多种形式的线上活动,鼓励全民宅家阅读、书香战“疫”。充分利用社交媒体的便捷性和交互性,引领全民宅家阅读。在疫情期间,各省级图书馆积极履行社会责任,通过征集活动向大众收集反映抗“疫”工作的作品,将具有历史价值与爱国教育意义的文献资料进行保存,传承新时代的奉献精神。

5.1.4 以青少年阅读推广内容为重点

在省级公共图书馆涉及的阅读推广对象中,儿童或者青少年是阅读推广的主要目标群体,在阅读推广活动中以少儿或青少年为主体的阅读推广活动较多,例如少儿英文、少儿作品征集、少儿诗词朗诵、少儿编程、青少年经典导读、青少年读书会等,青少年阅读是全民阅读的基础和未来,应当在培养青少年良好阅读习惯的过程中积极发挥父母的引领作用。图书馆在这过程中积极地开展了有关的家庭亲子阅读活动,例如:亲子沟通分享课、亲子绘本阅读示范课、亲子共读等,有利于培养良好的亲子关系和良好的阅读习惯。

5.1.5 积极推广数字阅读服务

各省级公共图书馆通过丰富数字资源来为读者提供便捷的资源获取和利用服务,并积极创新服务方式,利用新型社交媒体扩展图书馆服务,引导并推动全面阅读向传统阅读与数字阅读相结合的方向转变。尤其是在疫情时期以数字资源为主题的阅读推广内容占据了多数,图书馆通过打卡读书、扫描二维码、数字图书馆、数字资源链接、数字展厅、数字文化培训班、数据库推荐等形式向公众开展数字阅读服务,加强数字资源宣传与线上服务,培养用户的数字阅读素养,推进数字化阅读服务。

5.2 阅读推广内容主题演化特征

5.2.1 注重依据时事热点变化而调整推送内容

在2020年1月23日到2020年4月8日期间,全社会对Topic5(疫情防护与健康)的关注程度空前居高。Topic1(家庭与亲子教育)、Topic2(生活、情感)、Topic4(传统文化)、Topic6(诗歌艺术鉴赏)、Topic7(爱国教育主题)和Topic8(经济、政治、科技、哲学)等主题因为与疫情防护与健康的相关性不大。省级公共图书馆纷纷减少与Topic1、Topic2、Topic4、Topic6、Topic7和Topic8等主题相关的阅读推广信息发布数量,而增加

了与 Topic5 相关的阅读推广信息发布量。因而 Topic1、Topic2、Topic4、Topic6、Topic7 和 Topic8 等主题的阅读推广信息发布数量在 2 月份均呈现明显的下降趋势。当疫情缓和,社会生产生活秩序恢复正常后,这些主题下的阅读推广信息发布工作才开始回暖。Topic0(数字资源推广与获取)所讨论内容有助于人们保持社交距离和维持学习、工作和生产秩序,同时,Topic3(科学普及)所讨论内容能够普及新型冠状病毒肺炎的防护知识、预防措施、治疗方案等内容。因而省级公共图书馆为了向公众科普疫情知识与科学健康的防护措施,在 1 月至 4 月期间,一方面,通过推送防护知识、科普防疫知识等方式为公众解答疑惑并对读者群体进行心理调节,另一方面运用新媒体技术强化线上阅读推广活动,通过共享数字阅读资源鼓励读者宅家阅读,共同战“疫”。如浙江图书馆推出“21 天阅读|阅读静心 书香战‘疫’”,15 本热门电子书免费阅读,旨在激励用户做好自身防护并为国家的抗疫战贡献力量;陕西省图书馆通过征集抗击新冠肺炎疫情文献资料、AR 创意摄影抗击疫情主题大赛等活动记录抗疫过程中的感人瞬间并向用户传递疫情防控的正能量。Topic0 与 Topic3 的信息推送发布数量在此期间呈现一定的上升趋势。Topic7(爱国教育主题)反映出政府组织全社会力量在奋力抗疫,并尽最大努力保障人民生命财产安全,因而相关爱国、感动、感激成为人们探讨的特点,省级公共图书馆对相关阅读推广信息的推送数量略有增加。这进一步体现出图书馆在突发公共事件中承担社会责任并积极发挥图书馆的社会职能。

5.2.2 以纪念日、节日等为内容推送契机

由图 7 可知,阅读推广主题的峰值出现分别与“世界读书日”“元旦”“春节”“端午节”“建党日”等纪念日、节日的时间点相契合。说明省级公共图书馆在推送内容的选择方面考虑到了节假日的因素。究其原因,一方面,宣传纪念日、节假日等特殊的日子,让读者形成良好的历史观、节日观,了解历史文化和民俗传统以增强民族自豪感,凸显公共图书馆的使命。另一方面,借助于特殊的日子进行阅读推广活动,更能争取到相应推广经费,形成线上与线下的系列互动推广活动,活动影响力更大。

5.3 阅读推广内容主题信息发布数与阅读量分布关系特征

由图 6 和表 2 可知,每个主题类别对应的阅读推广信息发布数量的大小排序与每个主题类别下信息的阅读数量排序大致一致,公共图书馆发布的信息基本满足了读者的信息需求。由图 6 可知,每个主题类别对应的阅读推广信息发布数量的大小排序为 Topic2、Topic0、Topic8、Topic5、Topic1、Topic4、Topic6、Topic3、Topic7。由表 2 可知,依据平均阅读量大小对各主题排序依次为:Topic2、Topic5、Topic8、Topic1、Topic0、Topic4、Topic3、Topic6、Topic7。由此,可知,Topic2(生活、情感)为最受读者欢迎和最受图书馆员青睐的话题,其信息发布数量排序与阅读数量排序相匹配。Topic2、Topic0、Topic8、Topic5 和 Topic1 在这两种统计数据中都处于第一方队,而 Topic4、Topic6、Topic3、Topic7 在这两种统计数据中都处于第二方队。

6 研究启示

6.1 加强对微信阅读推广内容的科学规划配比

各省级公共图书馆应该做好推广内容的规划,推广内容要有计划、有目标、有选择、有针对、有质量,而不能盲目、随机、临时地去安排微信推广内容。微信阅读推广推文内容的选择应该重点协调以下关系:①九大主题的比例应协调,应依据公共图书馆的馆藏特色、读者需求和图书馆的主要职责进行设置;②旧书推荐与新书推荐的比例适中,既要考虑对读者保持新鲜感与吸引力,又要传承典籍与宣传中国传统文化,还要提升外借率和阅览量不高的图书等文献的利用率;③知识推荐与文献整体推荐的比例相促进,既要浅显地介绍知识让读者通过阅读微信能学到知识,又要引导读者去翻阅文献,进行深入阅读;④时事热点与常规推荐的配比恰当,既要体现时代担当和公共图书馆的责任感,并通过热点进行二次创作引导读者阅读相关书籍并使用相关服务,又要凸显公共图书馆的资源与服务特色并一以贯之。此外,还要解决好内容的体系化、深入化的要求与碎片化的形式的矛盾,可以采用“系列”“专题”等方式将一个主题内容分成多次发送,既解决了长篇大论带来的枯燥无趣,又可循循善诱,引导读者慢慢从浅入深、从零散到完整地去阅读了解一种文献或者服务。在前述分析结果的基础上,设置评判指标体系,结合微信阅读推广活动的总体得分和读者的反馈意见研判主题内容选择推送的合理性。

6.2 微信公众号阅读推广内容的从属地位在居家隔离期间的逆转

以往,阅读推广的主阵地在于线下活动。而疫情爆发的居家隔离期间,以微信公众号的方式进行阅读推广,可以让人民保持社交距离,达到疫情防控效果的同时有事可做。此时,以微信公众号方式进行阅读推广活动的重要性凸显。公共图书馆需要适应这种转变,要有以微信公众号为主阵地的思维,在 Topic0(数字资源推广与获取)的基础上精心筛选优秀的电子书

籍、电子阅读 APP、电子阅读数据库进行推荐，还要加大推荐的力度和深度。将一些原先的线下活动内容想办法以线上内容的形式进行展示，以填补线下不能进行的空白，让人们被线上内容吸引，不自觉地更喜欢以微信公众号为代表的线上推广方式。相关运营馆员要掌握一定的媒体工作知识背景，同时要具备较强的文字功底，只有以专业的知识素养作为支撑，才能合理并创新运营微信公众号，正确并有创造性地把握微信推文内容的质量，使微信公众号真正成为进行阅读推广服务的强大助力。

6.3　针对特殊人群的阅读推广内容的设置需加强

通过以上讨论发现，当前开展的阅读推广主题所面向的人群主要涵盖了幼儿、青少年、普通人群，但当前省级公共图书馆所面向的服务人群中针对特殊人群的阅读推广内容较少，国际图联图书馆特殊人群服务委员会对特殊人群的定义是“不能正常使用图书馆资源的人”[20]，会重点关注“因生活条件或身体、精神与认知障碍无法使用现有图书馆服务的人。这些人包括但不限于下列人群：在医院或监狱的人、无家可归的人、在养老院和其他保健设施的人。”[20] 特殊人群接受图书馆服务本身具有一定的困难性，但是基于微信平台的阅读推广内容具有独特的非到馆性、全民性、开放性等特性，能够打破线下服务的诸多局限，特殊人群更易接受此类型的服务。因此图书馆应该注意利用微信推广内容为特殊人群提供服务，酌情增加面向特殊人群的推广内容。在公众号关于特殊人群的推广内容设置中需要有专业的馆员进行设计，充分考虑特殊人群的群体特性，让公众号提供具有针对性的特殊的推广内容。

6.4　推广内容及其呈现形式创新设计

微信公众号的内容运营遵循着“引起注意——产生兴趣——主动搜索——开始行动——体验分享”的路径[21]。不同内容因其自身的特征属性受到用户的关注程度不同，例如社会热点内容更加容易吸引读者的关注，而经典文学类的内容受到读者关注的程度相对于社会热点内容来说其范围要小，这就需要推广主体对不同的阅读推广内容进行创新设计，解构加工，并灵活运用网络素材以及风趣幽默的语言风格来拉近与用户的距离，兼用记叙、议论、抒情等多种表达方式调动用户有意识地对阅读推广信息的加工和认知，符合现代移动阅读特征的写作方式，例如，语言表达简单化、口语化，突出设计标题引人注目的特点，在内容编辑方面综合运用图片、视频等突出重点内容。还可以引入视频流的方式进行“说书”或者“讲书”的形式，吸引读者。除此之外，对目标读者人群进行细分、锁定，能够更有针对性地设计满足读者需求的内容。要以这些创新的内容和形式引导读者进行深入阅读和思考，注重读者阅读质量的提升。

6.5　打造品牌内容

公共图书馆阅读推广服务在兼顾社会各类人群需求的同时，也应该结合本馆自身的服务特点，打造特色内容、品牌内容。阅读推广的品牌内容是图书馆阅读推广服务体系的最佳表现方式之一。它不仅能够体现图书馆的服务优势，彰显图书馆的服务个性和理念，更能够吸引读者的兴趣，形成品牌效应，推动阅读推广内容被更多的人了解和接受。基于微信公众号平台的推广内容更需要形成品牌链。由于微信公众号的特点和功能所决定，相较于线下服务的方式，读者更易于从微信推文内容中获取往期的推文内容，更容易参与到阅读推广的服务中去。图书馆可以根据自身的特点，参考九大主题的演进特点以及用户的接受程度，确定本馆的品牌内容，在推文内容的建立中做到既有普适性，又有特殊性。

7　不足与展望

本文以定量方法，统观全局，对 31 个省级公共图书馆的微信平台阅读推广内容主题归类情况及相关特征进行了验证和分析，比经验化的分析更加直观、准确。本文研究也存在不足：首先，在数据采集方面，由于数据采集难度较大，仅获取了 9 个月的数据，周期较短。其次，数据分析视角较为单一，囿于 LDA 主题模型，仅对 31 个省级公共图书馆的微信平台阅读推广内容主题方面进行了探讨。再者，仅进行了 31 个省级公共图书馆的微信平台总体数据分析，由于篇幅有限，没有以单个图书馆为单位进行数据分析。最后，计算机识别词语是一个机器学习的过程，文档的主题分类达不到完全准确，存在一定的局限性。

在未来，可以进一步开展的研究有：①可以跟踪多个年份的数据，对这些数据开展研究，例如获取 5 年时间的数据进行分析。②应该结合推广效果分析等多维度、多视角、综合立体化地分析这些阅读推广信息，得出的结论会更加立体丰满。③在此基础上，可以先分析每个公共图书馆的情况，再综合对比多个图书馆的状况。

参考文献：

[1] 周雨萌. 腾讯发布 2021 年第三季度财报[N]. 深圳特区报, 2021-11-11(A04).

[2] 李武,王丹,黄丹俞,等. 图书馆阅读推广研究十年进展(2005-2015)[J]. 图书馆论坛, 2016(12): 54-65.

[3] 张怀涛. 阅读推广的概念与实施[J]. 河南图书馆学刊, 2015(1):2 – 5.
[4] 王波. 阅读推广、图书馆阅读推广的定义——兼论如何认识和学习图书馆时尚阅读推广案例[J]. 图书馆论坛, 2015(10):1 – 7.
[5] 万慕晨,欧亮. 基于微信公众平台的高校图书馆阅读推广效果实证研究[J]. 图书情报工作, 2015, 59(22): 72 – 78.
[6] 蔡丽萍,孔德超. 基于 WCI 的省级公共图书馆微信阅读推广研究[J]. 图书馆工作与研究, 2016(10): 90 – 95.
[7] 王宝英,王宝玲,靳月庆. 静态阅读推广资源视角下的高校图书馆微信公众号探究[J]. 图书馆论坛, 2018, 38(4): 111 – 118.
[8] 中国政府网. 中华人民共和国行政区划[EB/OL]. [2021 – 10 – 01]. http://www.gov.cn/guoqing/2005-09/13/content_5043917.
[9] 梁晨晨,李仁杰. 综合 LDA 与特征维度的丽江古城意象感知分析[J]. 地理科学进展, 2020(4): 614 – 626.
[10] 王雅茹. 基于云南旅游在线网站吐槽数据的文本挖掘[D]. 昆明:云南大学, 2019.
[11] BLEI D M, NG A Y, JORDAN M I. Latent Dirichlet Allocation [J]. Journal of machine learning research, 2003(3): 993 – 1022.
[12] 邱均平, 沈超. 基于 LDA 模型的国内大数据研究热点主题分析[J]. 现代情报, 2021, 41(9): 22 – 31.
[13] CSDN 社区. LDA 困惑度 perplexity 的一些个人理解[EB/OL]. [2021 – 12 – 06]. https://www.csdn.net/tags/NtTaYgzsNzY0NjQtYmxvZw0000000000.html.
[14] 浙江图书馆. 浙图锦鲤 2.0 全面升级! 海量书籍、Kindle 阅读神器、旅行套餐……带上诗书去远行[EB/OL]. [2021 – 10 – 01]. https://mp.weixin.qq.com/s/gLcE6qFa-gdGkpXhYuP-wmw.
[15] 大圣的小宇宙. 如果不是因为一场意外,大唐诗坛一哥或许不是李白[EB/OL]. [2021 – 10 – 01]. https://mp.weixin.qq.com/s/IGTFVpBaXqaAn99Usenn7w.
[16] 艺非凡. 中国南海首次发现鲸落:一鲸落,万物生! 这是世界上最浪漫的死亡……[EB/OL]. [2021 – 10 – 01]. https://mp.weixin.qq.com/s/Q66gm4Vraag9O15EQOYZWg.
[17] 大才子造了一个谣,骗了我们几百年[EB/OL]. [2021 – 10 – 01]. https://mp.weixin.qq.com/s/vzH01LzpEnF4bHL51jkiBQ.
[18] 西藏自治区图书馆. “四讲四爱”激发西藏脱贫攻坚动能[EB/OL]. [2021 – 10 – 01]. https://mp.weixin.qq.com/s/6exLFNvJadMJqC16qKOJtg.
[19] 吉林省图书馆. 一“荐”倾心 |《读懂古典诗词》[EB/OL]. [2021 – 10 – 01]. https://mp.weixin.qq.com/s/vcBF-tL17GiK8uJOuVWiRw.
[20] 范并思. 阅读推广与图书馆学:基础理论问题分析[J]. 中国图书馆学报 ,2014, 40(5): 4 – 13.
[21] 崔永欣. 一流大学图书馆微信阅读推广内容营销策略研究[D]. 天津:天津工业大学, 2020.

作者贡献说明:
李倩:选题、整体结构确定,论文修改;
王帅:文献调研与整理,论文构思,论文起草。

A Study on the Topic of WeChat Platform's Reading Promotion Contents in Public Libraries by Using LDA Model in China

Li Qian[1,2] Wang Shuai[2]
[1] School of Economics and Management, Shanxi University, Taiyuan 030006
[2] Institute of Scientific and Technical Information of China, Beijing 100038

Abstract: [**Purpose/Significance**] This paper aims to identify the content themes and evolutionary characteristics of reading promotion by using WeChat public platform in domestic public libraries. [**Method/Process**] This paper defined and selected the reading promotion contents by obtaining the tweets of 31 provincial public libraries on WeChat in China, and used the LDA theme model to realize the theme recognition, and analysis of evolution and reading times of reading promotion contents. [**Result/Conclusion**] It is found that the topics of reading promotion carried out by public libraries with WeChat can be divided into nine categories, covering a wide range, each topic shows different evolutionary trends over time and the amount of reading for each topic also reflects a different degree of popularity. Finally, from the perspectives of scientific planning and matching of the promotion contents, the reversal of the subordinate status of the promotion contents, the promotion content setting of special groups, the innovative design of the promotion contents and its forms, and the creation of brand promotion contents, suggestions on the future WeChat reading promotion work of public libraries are provided.

Keywords: LDA model WeChat public platform reading promotion theme mining